

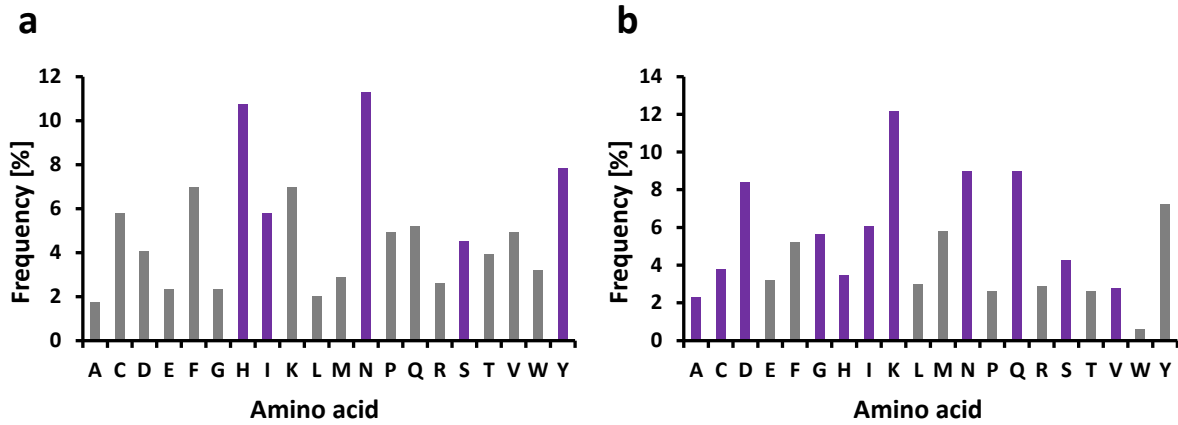
Supplementary information for:

Optimized tuning of TALEN specificity using non-conventional RVDs

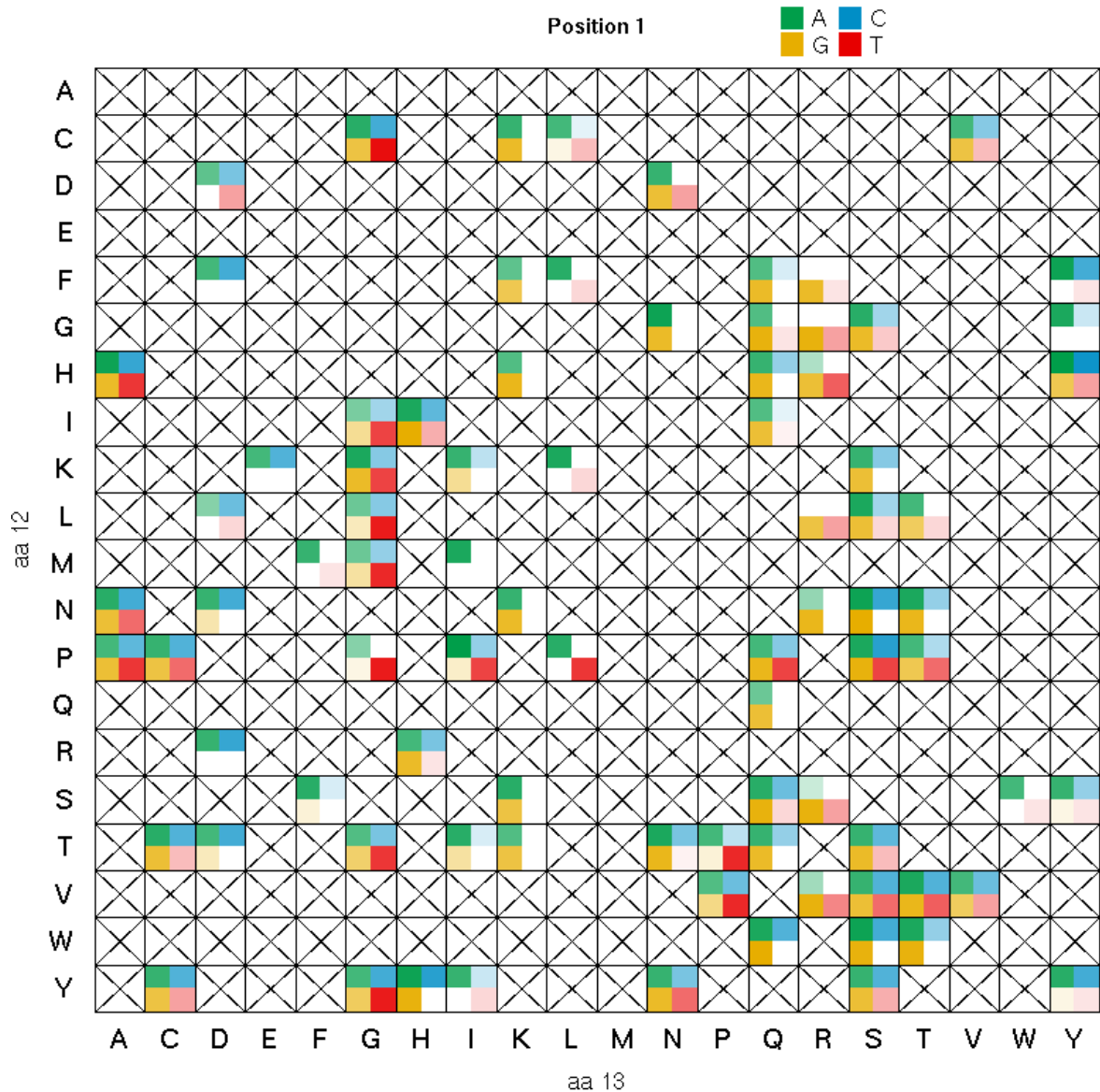
Alexandre Juillerat¹, Coline Pessereau¹, Gwendoline Dubois¹, Valérie Guyot¹, Alan Maréchal¹, Julien Valton¹, Fayza Daboussi¹, Laurent Poirot¹, Aymeric Duclert¹ and Philippe Duchateau¹

¹CELLECTIS S.A., 8 Rue de la Croix Jarry, 75013 Paris, France.

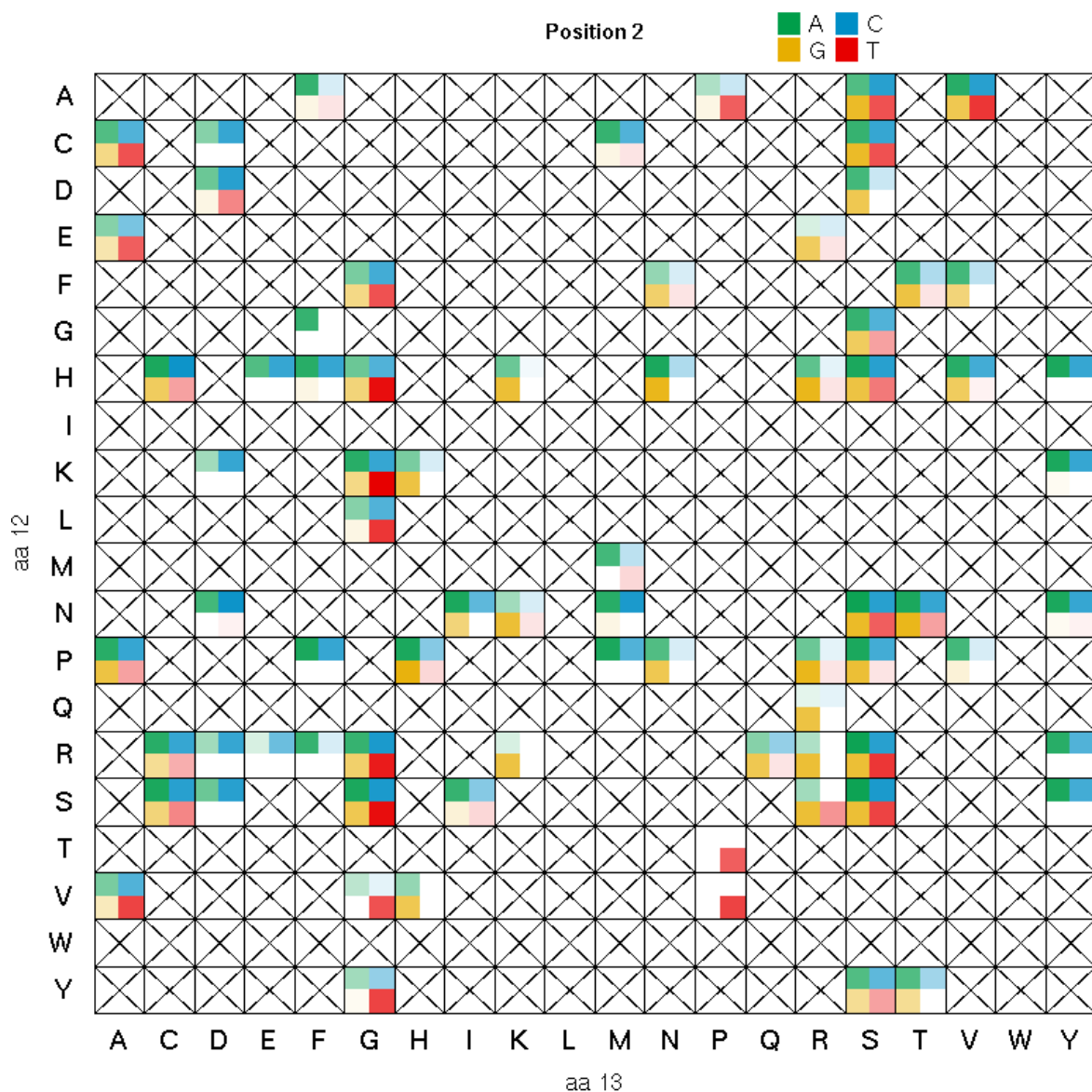
Corresponding authors: alexander.juillerat@cellectis.com and
philippe.duchateau@cellectis.com



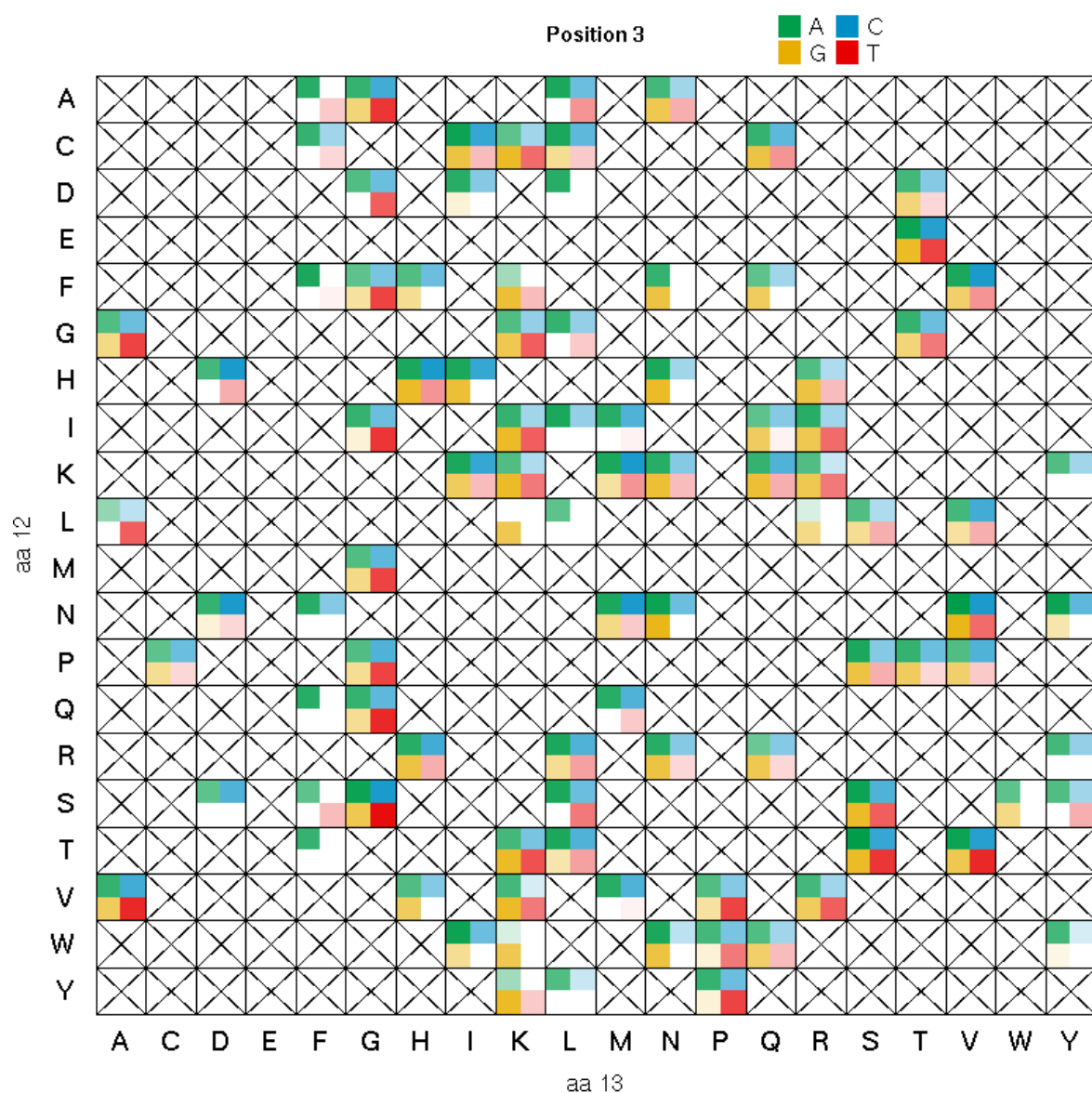
Supplementary Fig. 1 | Frequency of apparition of each amino acid identified in the yeast nuclease activity assay. (a) Amino acids found at position 12. (b) Amino acids found at position 13. Naturally occurring amino acids are indicated in violet. The frequency was normalized to the codon degeneracy for each amino acid.



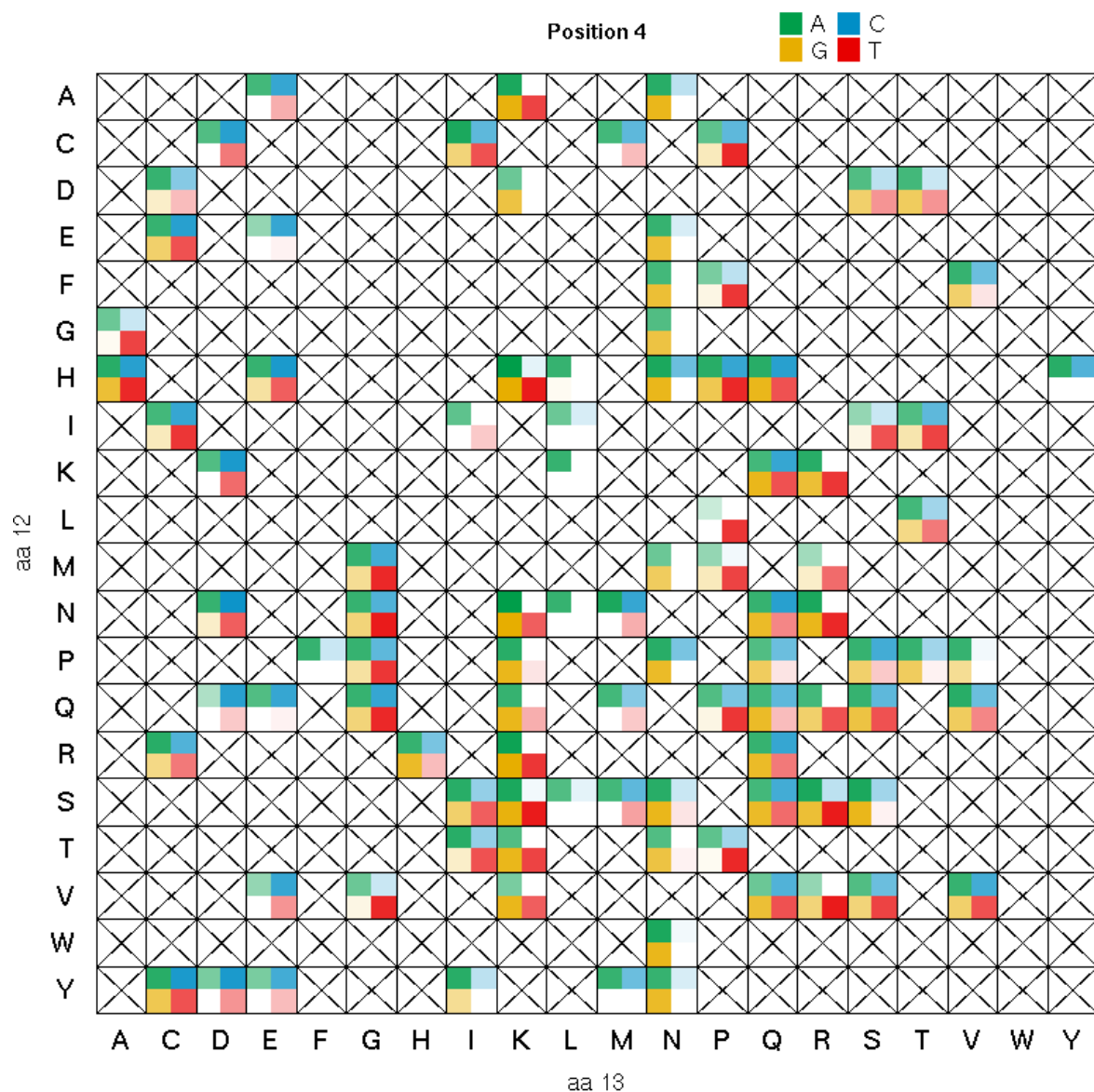
Supplementary Fig. 2 | Analysis of nuclease activity and specificity in function of amino acid 12 and 13 identity in position 1 of the array. Heatmap showing the nuclease activity of identified combinations of amino acids at position 12 and 13 of an RVD on targets containing the four possible nucleotides. A cross indicates that the combination of amino acids was not found in the present analysis of TALE array libraries. Activity is represented on a scale from 0 to 1 (arbitrary units).



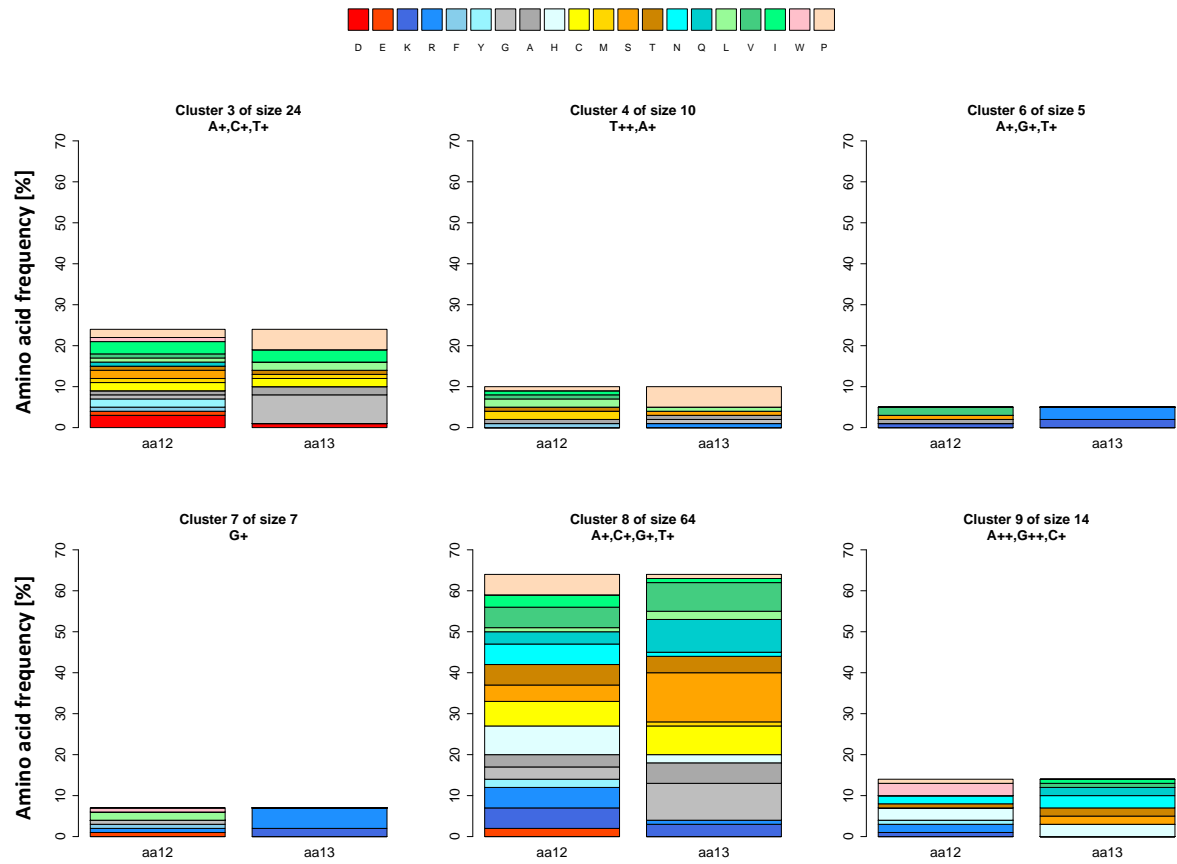
Supplementary Fig. 3 | Analysis of nuclease activity and specificity in function of amino acid 12 and 13 identity in position 2 of the array. Heatmap showing the nuclease activity of identified combinations of amino acids at position 12 and 13 of an RVD on targets containing the four possible nucleotides. A cross indicates that the combination of amino acids was not found in the present analysis of TALE array libraries. Activity is represented on a scale from 0 to 1 (arbitrary units).



Supplementary Fig. 4 | Analysis of nuclease activity and specificity in function of amino acid 12 and 13 identity in position 3 of the array. Heatmap showing the nuclease activity of identified combinations of amino acids at position 12 and 13 of an RVD on targets containing the four possible nucleotides. A cross indicates that the combination of amino acids was not found in the present analysis of TALE array libraries. Activity is represented on a scale from 0 to 1 (arbitrary units).



Supplementary Fig. 5 | Analysis of nuclease activity and specificity in function of amino acid 12 and 13 identity in position 1 of the array. Heatmap showing the nuclease activity of identified combinations of amino acids at position 12 and 13 of an RVD on targets containing the four possible nucleotides. A cross indicates that the combination of amino acids was not found in the present analysis of TALE array libraries. Activity is represented on a scale from 0 to 1 (arbitrary units).



Supplementary Fig. 6 | Frequency of every amino acid in clusters 3, 4, 6, 7, 8 and 9. Each amino acid at position 12 and 13 is represented by a different color. The size of the cluster (number of different RVDs) and the targeting preference is indicated.

In-site	Off-site	dRVD1 (discr. score)	dRVD2 (discr. score)	dRVD3 (discr. score)
A	C,G	KL (0.81)	AF (0.72)	YI (0.56)
	C,T	GN (0.82)	KL (0.71)	HK (0.46)
	G,T	HY (0.65)	NM (0.65)	ND (0.54)
C	A,G	<i>QD (0.5)</i>	<i>RE (0.43)</i>	<i>EE (0.34)</i>
	A,T	<i>QD (0.5)</i>	<i>RE (0.43)</i>	<i>EE (0.34)</i>
	G,T	ND (0.65)	NM (0.61)	HY (0.54)
G	A,C	GR (0.93)	FR (0.81)	<i>LK (0.7)</i>
	A,T	<i>LK (0.7)</i>	FR (0.68)	GR (0.55)
	C,T	GN (0.76)	NK (0.61)	HK (0.59)
T	A,C	LP (0.56)	<i>GR (0.38)</i>	<i>MP (0.34)</i>
	A,G	LP (0.56)	<i>MP (0.34)</i>	<i>VG (0.33)</i>
	C,G	PL (0.8)	<i>VG (0.59)</i>	<i>TP (0.54)</i>

In-site	Off-site	dRVD1 (discr. score)	dRVD2 (discr. score)	dRVD3 (discr. score)
A	C,G,T	KL (0.71)	AF (0.6)	YI (0.56)
C	A,G,T	<i>QD (0.5)</i>	<i>RE (0.43)</i>	<i>EE (0.34)</i>
G	A,C,T	<i>LK (0.7)</i>	FR (0.68)	<i>GR (0.55)</i>
T	A,C,G	LP (0.56)	<i>MP (0.34)</i>	<i>VG (0.33)</i>

Supplementary Fig. 7 | Potential non-conventional discriminating RVDs. Potential non-conventional discriminating RVDs (ncRVDs), identified in the library screens that presented discrimination between multiple nucleotides. These ncRVDs were determined especially by maximizing the difference of nuclease activity on the different nucleotides (favouring a strong nuclease activity on the desired base and a weak or no nuclease activity on the other bases). Italic indicates either non negligible residual activity on the weaker nucleotide, or an intermediate activity on the stronger nucleotide. Value in parenthesis represents the discrimination score (scale from 0, no discrimination to 1, maximum discrimination, arbitrary units).